TERESA MARQUÉS-AGUADO
(Universidad de Malaga)

Lexical Expressions of Possibility in Specialized Discourse: a Corpus-based Analysis[1]

In the last decades, there has been a growing interest in how specialized discourse is constructed (Hyland 1998, Gotti and Dossena 2001). In spite of the objectivity which is usually attributed to specialized discourse, subjectivity eventually surfaces when using modality, whereby authors take a stance on what is being argued, thus revealing their commitment to the factuality of the statements exposed.
Accordingly, the aim of the present article is to discuss the use of possibility (one of the values included within modality) in specialized discourse by resorting to a corpus of research articles compiled from several journals devoted to the field of Engineering. This corpus of written texts, which is a sizeable sample of more than three million words, will allow us to analyse how several lexical devices (that is, nominal, adjectival, adverbial and verbal) other than modal verbs are used to express possibility. The frequency of these expressions in the abstracts and the articles themselves will be also examined, highlighting any possible difference observed between them.

1. Introduction

Scientific discourse has been traditionally viewed as a register characterized by its objectivity and its neutral language, where the presence of the author is therefore diluted. Yet, several scholars, such as Swales (1990) or Hyland (1998), have recently looked into the ways in which the authors' commitment transpires in their writings, as well as the objectives they try to attain by doing so. By using these resources,

[1] The present research has been supported by the Ministry of Education (FPU programme, grant number AP-2004-244), and by the research projects granted by the Ministry of Science and Innovation (FFI2008-02336/FILO, "Grupo Consolidado") and by the regional government of the Junta de Andalucía (Proyecto de Excelencia P07-HUM-02609). These grants are hereby gratefully acknowledged. I would also like to thank Melchor Serrano Blanca for giving me permission to use the raw data of his corpus of Mechanical Engineering texts (end-of-year project, supervised by Prof. Antonio Miranda García) as a starting point for my research. An earlier draft of this paper was presented at the 28th ICAME Conference (University of Birmingham, May 2007).

authors try to mitigate statements in order to gain acceptance within the scientific community, thus avoiding categorical assertions of claims (Hyland 1996: 435). A key concept is that of *stance*, which has been defined as "the ways authors project themselves into their texts to communicate their relationship to subject matter and the readers" (quoted in Koutsantoni 2004: 163). In fact, the role of the readers proves to be central, as the author must develop means of persuading them and gaining their acceptance (Hyland 1996: 435; Swales 2004: 218). Thus, in Giltrow's words, "the language of science is seeded with social interest" (2005: 173), although specialists cannot forget either to keep the "neutral and unmarked tone (as this is a basic feature of specialized discourse)" (Gotti and Dossena 2001: 14).[2] In this social and communicative environment, modality "relate[s] to the interpersonal aspects of language" and surfaces when furthering an argumentation, indicating politeness, or showing the movement from observed data to generalizations (Vihla 2000: 600). It is, therefore, a signal of markedness (Gotti and Dossena 2001: 13).

From a linguistic point of view, authorial commitment can be perceived through their use of modality. In order to contribute to the recent research on modality, this article explores the value of epistemic possibility by way of linguistic expressions other than modal verbs in a corpus of scientific and technical research articles (henceforth, RA).[3] Accordingly, two major objectives are pursued: on the one hand, the use of different lexical items (nouns, adjectives, adverbs and lexical verbs) will be analysed; on the other, any differences arising between their occurrences in the articles and the abstracts (including the word-classes just mentioned) will be explored and evaluated.[4]

Therefore, the article is divided into the following sections: first, a brief theoretical account of modality; second, a description of the methodology followed; third, some general statistical data concerning

[2] Context also proves to be an essential variable in the development of modals, as shown by Fischer, as it "includes both the social and the communicative situation in which the utterances take place" (2003: 30).

[3] An analysis of the expressions of certainty found in one of the journals under scrutiny in this article, *Engineering Structures* (*ES*), can be found in Marqués Aguado (2008).

[4] For the sake of clarity, RA will be used to refer to the entire article and the corresponding abstract, while the body text of the RA will be referred to as article, and the abstract will keep this label. In turn, the complete set of RAs will be referred to as corpus.

the corpus; fourth, the analysis of the modal expressions found in the corpus (i.e. the study of the contexts where these expressions occur, arranged into word-classes); fifth, the comparison between the articles and the abstracts in terms of linguistic expressions of modality; and, finally, the conclusions drawn from the study.

2. *The expression of modality: definition and classification*

For Bybee and Fleischman, modality adds "a supplement or overlay of meaning to the most neutral semantic value of the proposition of an utterance, namely factual and declarative" (1995: 2), and for Quirk *et al*., it reflects "the speaker's judgement of the likelihood of the proposition it expresses being true" ([1985] 2003: 219). Whereas the traditional view focuses on modal verbs only, more recent studies delve into other expressions of modality, such as those by Kourilová (1994: 65) or Hyland (1998: 45). For instance, Quirk *et al*. pay attention to the so-called central modals, as well as to marginal modals, modal idioms, semi-auxiliaries and catenatives ([1985] 2003: 137). Similarly, Krug has recently analysed the so-called 'emerging English modals' (2000), such as *want to*.

Besides, modality can be conveyed by means of other lexical items, called 'lexical modals' by Huddleston and Pullum (2002: 173), i.e. nouns, adjectives, adverbs and lexical verbs. Gotti and Dossena also refer to them as "quite diverse items, the most common being verbal, adverbial, adjectival and nominal expressions, together with particles, clitics" (2001: 10-11).[5]

The taxonomies and the terminology used to refer to these modal expressions are, however, quite flexible and they vary depending on the theoretical approach followed. The traditional division into deontic modality (which involves obligation, permission and prohibition, i.e.

[5] Perkins acknowledges only partially the importance of such items when asserting that modal expressions in English "are used almost incidentally as paraphrases which serve to illuminate the meanings of the modal auxiliaries that are the primary focus of interest" (1983: 2). This is also the case of comprehensive reference works such as the *Oxford English Grammar* (Greenbaum 1996: 81) or the *Introduction to the Grammar of English* and *English Grammar: an Outline* (Huddleston 1984: 166 and 1988: 79, respectively).

"whether something is to be done" [Gotti and Dossena 2001: 11]), epistemic modality (which "has to do with the possibility or necessity of the truth of propositions" [Bybee and Fleischman 1995: 4]) and dynamic modality (which implies ability and willingness) goes back to von Wright's approach in the field of modal logic (1951: 1-2 in Hoye 2005: 1486). Some scholars have argued in favour of the existence of a fourth kind, evidential modality, whereby "instead of making a judgement about the truth-value of the proposition, the speaker offers evidence for it" (Palmer 2003: 7),[6] although there is no general consensus.[7]

3. Methodological procedure

This section, which explains the empirical procedure followed, is divided into two subsections: the compilation of the corpus (section 3.1.); and the retrieval of the data (section 3.2.). It must be added that from a methodological point of view corpora were rarely used before the 1980s when dealing with modality, as explained by Holmes (1988: 40). The reasons why RAs have been chosen as the object of analysis have been put forward by various authors: their pre-eminent position in research publications and their contribution to the dissemination and ratification of knowledge (Koutsantoni 2004: 164), the relationship they reveal between a discourse community (Hyland 1996: 452), and their "quantitative and qualitative pre-eminence" (Swales 1990: 93).

3.1. Compilation of the corpus

Three journals have been selected to compile the corpus, entitled *Engineering Structures* (hereafter *ES*), *Journal of Constructional Steel Research* (*JCSR*) and *Thin-Walled Structures* (*TWS*), all of them

[6] Martín Arrese's recent volume on evidentiality and modality (2004) contains several definitions of evidentiality, as shown in Pinar Sanz's review (2008). One such definition proposes that it is a "universal semantic dimension which regulates communication in general and the relationship established between the different participants in the communicative act" (Pinar Sanz 2008: 233). See also Ziegeler's discussion (2003: 45-49).

[7] These four kinds of modality have been grouped or re-formulated in various ways depending on the approach adopted, for instance: a) propositional and event modality (Palmer 2001: 8); b) epistemic and root modality (Coates 1983: 18-21; Gotti *et al*. 2002: 21); c) intrinsic and extrinsic (Greenbaum 1996: 80; Biber *et al*. 2000: 485); and d) epistemic, speaker-oriented and agent-oriented modality (Bybee 1985 in Bybee and Fleischman 1995: 6; Krug 2000: 42).

published by ELSEVIER, from which the RAs have been imported. These journals have been chosen on account of two factors: on the one hand, they belong to the same field (Mechanical Engineering); on the other, these publications are ranked among the top ones in their field.

ES has been taken as the basis of our study and, accordingly, most of the RAs come from this journal. The time-span under scrutiny ranges from 1995 to 2005, i.e. 10 volumes (numbers 17 to 27), which usually comprise 12 issues, although they can exceptionally contain from 10 to 14 issues. As for *JCSR* and *TWS*, only the volumes published in 2004 (volumes 60 and 42, respectively), with 12 issues each, have been selected. A total of c. 850 RAs has been downloaded, although some of them have been excluded, as they focus almost entirely on mathematical demonstrations, and so have been those sections dealing with formulae. The reason behind this is that *Wordsmith Tools* (Scott 1998), the programme used to retrieve the data, counts formulae and scientific notations (such as *F*) as words, thus distorting the results obtained. Moreover, they are not related to our study.

For the purpose of our analysis, the RAs have been downloaded into Word files and then they have been split into three smaller files: one including the abstract, another with the article, and the third one with the references, which have been left aside owing to their lack of relevance for the present study. The rationale behind this procedure is that this will allow for the comparison of modal expressions between articles and abstracts.

When downloading texts from web pages, some distortions and errors may occur, and they have been amended manually. This process of revision comprised changes such as deleting the remaining formulae and hyphenation, given that *Wordsmith Tools* assumes that hyphenated words are two independent items.

3.2. Data retrieval

The data have been retrieved with *Wordsmith Tools* and, more specifically, with the tools called *Wordlist* and *Concord*. The Word files mentioned above have been converted into plain text files and, from these, two wordlists per journal have been generated: one for the article, and one for the abstract. Once the relevant expressions of modality have

been selected from the wordlists, their concordances have been generated in order to study the context where they appear. This is a key step inasmuch as it is important to discard those occurrences which do not imply any shade of modality, as in the case of the noun *probability*, explained in 5.1. This means that not only a quantitative, but also a qualitative analysis, has been carried out.

4. General statistical data

The corpus compiled is assumed to be representative of specialized discourse insofar as it amounts to over three and a half million tokens in the articles, plus c. 130,000 in the abstracts, as shown in Table 1:

ARTICLES	ABSTRACTS	CORPUS
3,518,397	133,972	3,652,369

Table 1. Total of tokens in articles, abstracts and the whole corpus

The statistical data are organised as follows: first, a comparison to other standard, non-scientific corpora and a comparison of the data relating to the articles and the abstracts; and second, a sample selection that reveals the potentiality of this corpus as regards the study of other aspects related to modality (*harmonic combinations* or *modal synergies*,[8] deontic or evidential modality, etc.).

4.1. Reliability of the corpus: comparison in terms of most frequent words (MFW)

The list of the 20 MFW extracted from the corpus has been compared to those of other corpora with the aim of: first, checking the reliability of our corpus; and second, testing whether the MFW differ from one corpus to another, and whether this is linked to the scientific character of the corpus under scrutiny, as shown in Table 2:

[8] *Modal synergy*, also called *cumulative modality* (Hoye 1997: 19) or *harmonic combination* (Hyland 1998: 150), has been recently defined by Hoye as the combination of two or more modal expressions, an example being the collocations modal verb - adverb (2005: 1494). See also 4.2.

POSITION	CORPUS	COBUILD	WORD FREQUENCIES
1	The	The	The
2	Of	Of	Of
3	And	And	And
4	In	To	A
5	To	A	In
6	A	In	To (infinitive)
7	Is	That	It
8	For	I	Is
9	Are	It	To (preposition)
10	With	Was	Was
11	Be	Is	I
12	As	He	For
13	By	For	That
14	That	You	You
15	This	On	He
16	On	With	Be*
17	At	As	With
18	Figure	,	On
19	From	Be	By
20	Load	Had	At

Table 2. Comparison of the corpus to other corpora

When compared to the COBUILD MFW list (Sinclair 1991: 143), most of the function words coincide, although some variations as regards their positions are to be noted. The main difference lies in the use of personal pronouns (*I*, *you*, *he* and *it*) in the COBUILD list, which are not included in our corpus MFW list, as expected, since they reveal a certain degree of subjectivity. Similarly, the lexical words *figure* and *load* are corpus-specific and, hence, are not ranked in the COBUILD list among the first 20 positions. The same differences apply as regards the position of some function words and the use of personal pronouns when observing Leech *et al*.'s MFW list (2001: 120), based on the *BNC*

(*British National Corpus*).[9] When comparing the ten MFW of the articles with those of the abstracts, a full correspondence is observed, except for the prepositions *in* and *to*, and the determiner *a*, which swap their positions, as shown in Table 3:

POSITION	ARTICLES	ABSTRACTS
1	The	The
2	Of	Of
3	And	And
4	In	To
5	To	A
6	A	In
7	Is	Is
8	For	For
9	Are	Are
10	With	With

Table 3. Comparison of MFW of articles and abstracts

4.2. Versatility of the corpus

Being so large, the corpus can be approached from a variety of perspectives. For example, although deontic modality has been described as infrequent in scientific RAs (Gotti and Dossena 2001: 14), some instances are found, such as examples (1) and (2). The relevant words are marked in bold hereafter:

(1) This implies that all thermal strains which were added in the first iteration **must be removed** in successive iterations in order to ensure that equilibrium conditions are maintained (*ES*)

(2) Therefore the FE model **must be able** to recognise the possible failure modes (*JCSR*)

Similarly, even evidential modality is seen to occur, as in examples (3) and (4), whereby report or sensory evidence is supplied. The use of

[9] Data calculated per million words. The *BNC* website is at <http://www.natcorp.ox.ac.uk/>.

evidential modality is justified by the fact that a great deal of research stems from previous claims, supports others' opinions or implicitly refutes previous arguments put forward by peers:

(3) Both the model and the experiment **have shown that** a frame without bracing has very weak racking resistance (*JCSR*)

(4) Witz and Tan **have suggested that** torsional slippage is possible as well as axial slippage (*ES*)

By retrieving the concordances of the modals or of any of the items conveying modality, the way they combine can be also surveyed, a phenomenon called *harmonic combination*, as explained above, and as illustrated in (5) and (6):

(5) **Possibly** in future experiments it **may be necessary** to find ways and means of improving the sampling rate even beyond HZ as was done in this experiment (*ES*)

(6) The stress concentration effect **may be seen clearly** in figure (*JCSR*)

5. Analysis of possibility in the corpus

In the following sections, the linguistic expressions that entail possibility are dealt with and organised into word-classes: nouns, adjectives, adverbs and lexical verbs. In each word-class, some items have been selected in view of the referred literature, rather than reading through the RAs to manually detect the relevant items, a methodology used in smaller corpora, such as the one investigated by Pérez-Llantada Auría (2008: 132). Indeed, a word-by-word analysis of each of the RAs is an unapproachable task owing to the size of the corpus. Absolute figures are offered, rather than normalized frequencies, which are given in section 6, where the data for the articles and the abstracts are compared.

5.1. Nouns

The relevant literature has reported that nouns are the word-class that is employed least frequently to convey possibility (Holmes 1988:

35-36). Table 4 shows the frequency of nouns found in our corpus that express possibility:

NOUNS	FREQUENCY
Possibility	42
Likelihood	31
Probability	26
Doubt	11
Speculation	2
TOTAL	112

Table 4. Nouns conveying possibility

The noun *likelihood* is typically followed by *of*, which is followed in turn by a noun phrase, as in (7), or an *-ing* form, as shown in (8). Yet, this noun has not been counted when appearing in technical terms such as *likelihood function*, as they are not related to the expression of possibility, as explained in 3.2.:

(7) […] the value of the classification tree is in identifying a small subregion of feature space that corresponds to **a high likelihood of good design** (*TWS*)

(8) The **likelihood of being** suitable for monitoring ISUs is small (*ES*)

The nouns *possibility* and *probability* are the most frequent ones in quantitative terms, but they do not always imply modality, especially in the second case, given that this noun generally forms part of technical concepts, such as *failure probability* or *probability distribution*. As a matter of fact, only 26 out of the 399 occurrences found in the corpus convey epistemic modality, as in (9):

(9) The mutation operation is performed on certain individuals on a bit by bit basis, with **a certain probability of** mutation (*ES*)

As for *possibility*, only 42× actually express modality. When post-modified, this noun is occasionally followed by *that* (3×), and most commonly by an *of*-phrase containing either an *-ing* form (12×) or a

noun phrase (21×), as illustrated in (10). Examples (11) and (12) show that an intensifying adjective (Quirk *et al*. [1985] 2003: 429) can also modify this noun:

(10) A major advantage of GBT is **the possibility of considering** an arbitrary number of deformation modes (*TWS*)

(11) There was **a high possibility of collapse** due to strong earthquakes (*ES*)

(12) Furthermore **a strong possibility existed that** the soil near the base of the stack was contaminated by previous oil spills (*ES*)

5.2. Adjectives

Some of the adjectives surveyed have been taken from Miranda García *et al*. (2001). There are some adjectives which show a relatively low frequency of occurrence, as shown in Table 5:

ADJECTIVES	FREQUENCY
Probable	37
Plausible	9
Liable (to)	8
Doubtful	7
Debatable	4
Seeming	1
TOTAL	66

Table 5. Less common adjectives conveying possibility

They mostly occur in either attributive or predicative positions (40×), an example of which is provided in (13), but very rarely in so-called 'impersonal expressions', which typically "obscure the source of judgements" (Hyland 1998: 132). These constructions follow the pattern 'anticipatory *it* + verb 'to be' + adjective + that',[10] i.e. when followed by a *that*-clause (or *whether* in the case of *doubtful*), as in (14):

[10] This structure has been identified as a recurrent syntactic pattern to convey authorial stance by Pérez-Llantada Auría (2008: 136).

(13) […] **probable causes of collapse** of the scaffold support system were discussed (*ES*)

(14) Although this is unfortunate **it is plausible that** the wear process would have deteriorated during the monitoring period (*ES*)

The most frequent adjectives to convey possibility are given in Table 6:[11]

ADJECTIVES	FREQUENCY
Possible	722
Likely	377
Unlikely	64
TOTAL	1163

Table 6. Most frequent adjectives conveying possibility

Possible shares both epistemic and root meanings, as remarked by Hyland (1998: 131) and, as a matter of fact, only 722 out of the 1376 occurrences of this adjective actually convey epistemic meanings (paraphrased as 'possible that' in Hyland's view). Indeed, *to*-infinitives with *possible* usually express deontic meanings. In turn, and contrary to the set of adjectives in Table 5, *likely* and *unlikely* tend to be followed by a *to*-infinitive, and the subject is, accordingly, deprived of human volition, whereby authors distance themselves from the propositions uttered (Hyland 1998: 123), as shown in (15). *That*-clauses, as in (16), amount to a total of 101×:

(15) The specimen tests allowed for the **different conditions likely to arise** in the manufacture of concrete: cured, well compacted with a poker vibrator, well compacted by hand, and self-consolidating without any vibration (*TWS*)

(16) **It is therefore very unlikely that** the crack closed during the vibration tests, which would require that the maximum strain during the test exceeds the residual strain (*ES*)

[11] Adjectives such as *possible* and *likely* belong to the set of adjectives typically used by authors "for assessing entities, facts and processes or when emphasising their value or usefulness" (Pérez-Llantada Auría 2008: 134).

Irrespective of the structure that follows the adjective, the verb preceding it is commonly *to be* in the present tense (as in examples 14 and 16), as noted by Perkins (1983: 67).

The only adjective which is frequent in both attributive and predicative positions is *possible*, although it is used attributively 505×, which means 69.94% of the total of the occurrences with epistemic values. Yet, Hyland has suggested that "adjectives are used attributively with an apparently restricted, but unspecified, range of nouns" (1998: 133). One such noun in our corpus is *cause*, which is preceded by *possible* (19×), *probable* (13×), *likely* (7×) and *plausible* (2×), as shown in (17) and (18):

(17) […] their interest was in the statistical characterization of the types of falsework failures and the identification of **the most probable causes** (*ES*)

(18) To describe **possible causes** of collapse of scaffold systems during construction of high clearance concrete (*ES*)

Some adjectives that can potentially convey the value of impossibility, such as *dubious*, are not recorded in our corpus. As expected in RAs that focus on scientific achievements, emphasis is placed on possibility, rather than on impossibility, hence the low frequency of negative adjectives, as also shown in Marqués Aguado (2008: 382). The alternative is usually the insertion of a negative particle in the verb phrase, as in *does not seem realistic* (*ES*) or *elsewhere it was also not possible to obtain* (*ES*).

5.3. Adverbs

According to Hyland, this word-class comes second after lexical verbs when dealing with epistemic modality (1998: 134). Most of the adverbs searched for in this section, and given in Table 7, are listed in Quirk *et al*. under the heading 'content disjuncts', and more specifically, under the group expressing the degree of truth being conveyed ([1985] 2003: 615, 620-623):

ADVERBS	FREQUENCY
Probably	169
Possibly	100
Perhaps	79
Presumably	11
Likely	9
Seemingly	9
Maybe	4
Supposedly	3
TOTAL	384

Table 7. Adverbs expressing possibility

In broad terms, disjuncts "express an evaluation of what is being said", to the extent that they are identified "with the speaker's authority for, or comment on, the accompanying clause" (Quirk *et al*. [1985] 2003: 440).[12] These items (referred to as 'adjuncts' by Huddleston and Pullum [2002: 612], i.e. "elements traditionally regarded as modifying the verb") reflect varying degrees of commitment to the truth of the statement on the part of the author, depending on the strength of modality (quasi-strong, medium and weak [Huddleston and Pullum 2002: 768-770]).

The most outstanding feature of modal adverbs is their mobility when functioning as adverbials, as they are optional and "peripheral in clause structure" (Quirk *et al*. [1985] 2003: 440), as shown in examples (19)-(22):

(19) **Perhaps,** the most intriguing result was the observation of a "boundary layer" effect (*TWS*)

(20) **Possibly** on account of weak nonlinearities of the building, the measurements obtained with sinusoidal tests seemed more suited than those obtained by sweep tests for the application of identification techniques (*ES*)

[12] According to Quirk *et al*., these adverbials belong to the set of approximation elements ([1985] 2003: 485).

(21) As the calculation was carried out elastically, the ground stiffness **was presumably overestimated** (*ES*)

(22) The differences between the two analyses **were probably caused** by stiffness variations in modelling the front steel frame (*ES*)

They can occasionally behave as modifiers of adjectives (23), or of noun phrases (24):

(23) It has been shown that the structural intensity analysis can act as **a new, maybe more reliable criteria** for marine structural design (*TWS*)

(24) Fire following earthquake is a potential major agent of damage of **possibly holocaust proportions** for both the United States and Japan (*ES*)

These adverbs typically hedge the probable cause for a response, outcome, etc. (conveyed by means of conjunctions such as 'because', prepositions such as 'due to' or main verbs such as 'attribute' or 'cause') in order to lessen the impact of what is being put forward, as in (25) and (26). This coincides with the most frequent noun when dealing with adjectives in attributive position, *cause*, as explained in 5.2.:

(25) However the influence of strain rate on the fracture rupture strain cannot be conclusive **possibly due to** the difficulty of carrying out the test on material up to the point of rupture (*JCSR*)

(26) The large increase in torsional loading **is probably caused by** buffeting on the windward side by the fluctuating edge of the wake (*ES*)

As for the combination of adverbs and modals, as in (27), 15× have been found out of the 384 adverbs analysed in our corpus. These instances cannot be explained on the grounds of pleonastic reiteration, but are rather "a matter for modal concord or cumulative modality" (Hoye 1997: 19):

(27) The discrepancies between results **may possibly be attributed** to the very small magnitude of slip (*ES*)

In view of these results, we cannot but partially agree with Hyland's opinion that "items [referring to adverbs] expressing doubt are rare in the corpus" (1998: 137), since the total number of occurrences of the adverbs retrieved from our corpus amounts to only 384.

5.4. Lexical verbs

Most of the lexical verbs considered here have been taken from Hyland's section on epistemic judgement verbs, which "indicate there is some conjecture about the truth of a proposition" (1998: 120), such as *believe*, *presume*, *hope*, *suspect* and *speculate*, which are shown in Table 8:[13]

LEXICAL VERBS	FREQUENCY
Appear	498
Seem	346
Believe	140
Suppose	52
Presume	23
Hope	18
Suspect	12
Speculate	4
TOTAL	1093

Table 8. Lexical verbs conveying possibility

A brief comment will be made on one of the most frequent lexical verbs, *seem*. *Seem* is typically followed by: a) an adjective loaded with connotative values – positive ones more often than not – (*encouraging*, *(in)appropriate*, *more effective*, *advisable*, etc.), as in (28), (29) and (30); b) a *to*-infinitive, usually *to be* (*to be adequate*, *to be very promising*, etc.), as in (30), (31) and (32); or c) a *that*-clause, as in (33). Generally speaking, *seem* contributes to evaluate the application of the current research (28), to mark its relevance in relation to the limitations

[13] An account of the progressive grammaticalization of *be supposed to* is offered by Ziegeler (2003: 33-69).

of the previous research (29), to claim centrality of the findings and give significance to the results obtained (31), or to hedge the probable cause for a given fact (32), as already discussed in sections 5.2. and 5.3. In this line, it is also employed to lessen the impact of a negative opinion, as shown in (30) and (34):

(28) It has been shown previously that the Kardo electrode **seems the most suitable** in terms of overall connection (*JCSR*)

(29) […] **it seems fair to argue** that its "analytical character" is quite **debatable** (*ES*)

(30) […] such a method **seems to be impractical** (*ES*)

(31) Fukumoto' equation is most unique and **seems to predict** test results rather well (*ES*)

(32) However **it seems to be caused** by the difference of the introduced initial bolt axial forces and not caused by the shear force (*JCSR*)

(33) […] **it seems that** the theories do not properly allow for the bulging effect of the bearing (*ES*)

(34) […] if that is the case, the computation of a damage index with a postprocessor **seems redundant** (*ES*)

The main point under discussion here is the choice between the active and the passive voice. Only 207× of the passive have been found, as shown in (35), as opposed to 866× of the active, as in (36). According to Hyland (1998: 122-123), passives are far less frequent when dealing with epistemic lexical verbs, and this conclusion also holds true in the corpus under study. In order to distance themselves from the point they are making and to appear more objective, authors typically use grammatical subjects that lack human volition, as in the cases of *presume*, *seem* or *appear*, as shown in (37); otherwise, verbs such as *believe* usually take a human subject (*we*, *the authors*, *many*) (38):

(35) **It is usually supposed that** the polarization direction of the PZT is the thickness direction axis (*TWS*)

(36) **The authors hope that** this paper will act as a catalyst, sparking interest and further research in these areas (*ES*)

(37) The order of **the following discussion does not presume** any relative importance; the relative order could only be judged for a specific case (*ES*)

(38) Based on extensive testing in our lab on different types of specimens and different stress states, **we believe that** the weighting function has three branches, depending on the activation of different failure mechanism (*TWS*)

The presence of *harmonic combinations* with modal verbs (60×), and mostly with the verb *may* (or, to a lesser extent, *could* (39)), must also be considered, and so must *harmonic combinations* with other lexical expressions indicating possibility (40):

(39) […] in the sweep test it **could seem** that there was **probably** a combination of the two modes (*ES*)

(40) […] though it **seems most likely** that the iron rods were a part of the original system (*ES*)

Finally, it should be noted again that either *that*-clauses, as exemplified in (33), (35), (36), (38) and (40), or *to*-infinitives, as illustrated in (29), (30), (31), (32) and (41), typically follow these verbs, either in the active or in the passive:

(41) The global optimization process begins with the identification of a set of features which, through expert judgment, **is believed to be** relevant to the optimization problem (*TWS*)

6. *Comparison between articles and abstracts*

Table 9 summarizes the absolute figures for the number of occurrences per word-class in the articles, in the abstracts, and in the corpus. Besides the absolute figures, the normalized frequencies (to a text of 100,000 words, see Biber 1998: 14) are also provided:

	NOUNS		ADJECTIVES		ADVERBS		VERBS		TOTAL	
	Abs.	**Norm.**	**Abs.**	**Norm.**	**Abs.**	**Norm.**	**Abs.**	**Norm.**	**Abs.**	**Norm.**
Articles	99	2.81	1184	33.65	380	18.80	1071	30.43	2734	77.70
Abstracts	13	9.70	45	33.58	4	2.98	22	16.42	84	62.69
Corpus	112	3.06	1229	33.64	384	10.51	1093	29.92	2818	77.15

Table 9. Absolute figures and normalized frequencies per word-class

These data reveal the following:

1. Lexical expressions of possibility in the abstracts are not as infrequent as could be conjectured from the absolute figures, as they amount to a total of 62.69. Although this figure is clearly below the ones for the corpus (77.15) and for the articles (77.70), the difference is not outstanding, and this allows us to conclude that these expressions are certainly regularly used in abstracts.
2. The four word-classes under scrutiny in the articles can be ranked in descending order on the basis of these normalized figures as follows: adjectives > verbs > adverbs > nouns. Yet, this sequence does not fully apply to the set of abstracts, where nouns show a higher rate than adverbs, in such a way that these word-classes are ranked as follows: adjectives > verbs > nouns > adverbs.
3. The normalized frequencies for each word-class show that the preference for particular lexical items is different for the abstracts (excepting adjectives), whereas the articles and the corpus behave in the same way. First, the frequencies of verbs in the corpus (29.92) and in the articles (30.43) almost double the one in the abstracts (16.42). Second, the rate of nouns in the abstracts clearly stands out if compared with the articles and the corpus, as it is three times as high. Third, adverbs are also approximately three times more frequent in the corpus than in the abstracts, and six times more frequent in the articles than in the abstracts, so that the differences become even more salient. As a matter of fact, the normalized figures for nouns in the articles and for adverbs in the abstracts are similar (2.81 and 2.98, respectively).

7. Conclusions

Several conclusions may be drawn from our study of possibility in the corpus compiled:

FIRST. The present study has focused on the various ways whereby the value of possibility is conveyed in the specialized discourse of Mechanical Engineering, and the results obtained allow us to conclude that RAs have proved to be valid to analyse modality. By and large, the main aim that authors pursue when using lexical expressions of possibility in these RAs is to publicise their claims by appealing to their readers and to convince them about the validity of their claims (as also suggested by Pérez-Llantada Auría [2008: 143]), but without imposing their views, in such a way that room is left for other options and opinions. Thus, such value has a clear communicative aim and an interpersonal function, as stated by Vihla (2000: 600).

SECOND. In terms of the absolute data retrieved for each word-class, despite the extensive range of modal expressions, only some of the items are used regularly in the corpus, such as *possible*, *probably* or *appear*, whereas others are employed only occasionally (e.g. *debatable*, *seemingly* or *speculate*). Moreover, certain syntactic environments seem to be favoured to convey possibility, such as the use of the active as opposed to the passive voice, the alternation between *to*-infinitives and *that*-clauses, the use of certain modal adjectives either predicatively or attributively, or the use of *harmonic combinations*. Yet, it must be noted that this is not an extensive analysis of all possible lexical items, because the size of the corpus has not allowed a manual and word-by-word check of the RAs. Thus, more items in each word-class can potentially be added.

THIRD. Several conclusions have also been derived from the comparison of the normalized frequencies provided for the abstracts, the articles and the corpus. First and foremost, there are not extremely marked differences in the total figures for lexical expressions of modality in the three groups, on account of the normalized figures obtained (as opposed to the absolute figures, which are misleading when comparing texts of different sizes). This means that while lexical expressions of possibility may be used in the article to further an argumentation or to hedge a tentative conclusion, they must be used for

an altogether different purpose in the abstracts, for instance to introduce the *likely* application of the research under way, to contradict the previous research by highlighting its *possible* shortcomings, or to suggest the *possible* ways of continuing the previous research, among other options. Nevertheless, they all coincide in that they indicate politeness, either towards other peers' previous research, or towards forthcoming amendments of one's own research. By the same token, and as noted in Sections 5.1. through 5.4., authors recurrently employ these lexical items to introduce the likely cause for a particular phenomenon or to create the niche for their own investigation.

Abstracts follow particular conventions in terms of the use of nouns, adverbs and lexical verbs, which are somewhat different from the ones in the articles. These conventions may derive from the space constraints that typically have an influence on abstracts, hence the extremely low frequency of adverbs and the high rate of nouns, for instance. If we delve into the particular normalized frequencies of nouns, adjectives, adverbs and lexical verbs, the present study has shown that although nouns have been reported to be the least frequent item when furthering possibility, this claim does not hold true in the case of abstracts. In the same vein, adverbs do not come second only after verbs (as claimed by Hyland [1998: 134]), but rather third and fourth (in the articles and the abstracts, respectively). Nonetheless, these conclusions should be further confirmed with other corpus-based studies.

FOURTH. From a methodological point of view, the qualitative analysis of the data retrieved has been of paramount importance, as it is necessary (though difficult) to establish when a judgement on the truth value is being put forward. For instance, *possibly* does not always convey epistemic possibility, but can also express deontic meanings. Similarly, verbs such as *suppose* in the imperative have also been ruled out, as they lack any shade of judgement. On the other hand, the validity of corpora for the study of modality has been corroborated with this analysis.

All in all, studies such as this one allow us not only to attain a better understanding of the mechanisms by means of which possibility is expressed in specialized discourse, but also to contribute to the field of ESP, so that teachers and textbooks can cater for their students' linguistic needs in depth, rather than by simply resorting to modal

verbs. Nevertheless, further research should broaden the scope by extending the study already carried out to the study of epistemic modality as a whole. Likewise, a comparison of the use of these lexical expressions of modality with the use of modal verbs themselves is also much needed in order to evaluate which of the two predominate, and whether any salient differences of use emerge.

References

Biber, Douglas, 1998, *Variation across Speech and Writing*, Cambridge, Cambridge University Press.

Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward, 2000, *Longman Grammar of Spoken and Written English*, Harlow, Longman.

The British National Corpus, version 3 (BNC XML Edition), 2007, Distributed by Oxford University Computing Services on behalf of the BNC Consortium, <http://www.natcorp.ox.ac.uk/>.

Bybee, Joan / Fleischman, Suzanne, 1995, *Modality in Grammar and Discourse*, Amsterdam / Philadelphia, John Benjamins.

Coates, Jennifer, 1983, *The Semantics of the Modal Auxiliaries*, London, Croom Helm.

Fischer, Olga, 2003, "The Development of the Modals in English: Radical Versus Gradual Changes". In: Hart, David (ed.), *English Modality in Context. Diachronic Perspectives*, Bern, Peter Lang: 17-32.

Giltrow, Janet, 2005, "Modern conscience: Modalities of obligation in research genres". *Text* 25/2: 171-199.

Gotti, Maurizio / Dossena, Marina, 2001, "Introduction". In: Gotti, Maurizio / Dossena, Marina (eds.), *Modality in Specialized Texts: Selected Papers of the 1st CERLIS Conference*, Bern, Peter Lang: 9-18.

Gotti, Maurizio / Dossena, Marina / Dury, Richard / Facchinetti, Roberta / Lima, Maria, 2002, "Introduction". In Gotti, Maurizio / Dossena, Marina / Dury, Richard / Facchinetti, Roberta / Lima, Maria (eds.), *Variation in Central*

Modals: a Repertoire of Forms and Types of Usage in Middle English and Early Modern English, Bern, Peter Lang: 19-44.

Greenbaum, Sidney, 1996, *The Oxford English Grammar*, Oxford, Oxford University Press.

Holmes, Janet, 1988, "Doubt and Certainty in ESL Textbooks". *Applied Linguistics* 9/1: 21-44.

Hoye, Leo Francis, 1997, *Adverbs and Modality in English*. London / New York, Longman.

Hoye, Leo Francis, 2005, "'You may think that; I couldn't possibly comment!' Modality Studies: Contemporary Research and Future Directions. Part II". *Journal of Pragmatics* 37: 1481-1506.

Huddleston, Rodney, 1984, *Introduction to the Grammar of English*, Cambridge, Cambridge University Press.

Huddleston, Rodney, 1988, *English Grammar: an Outline*, Cambridge, Cambridge University Press.

Huddleston, Rodney / Pullum, Geoffrey K., 2002, *The Cambridge Grammar of the English Language*, Cambridge, Cambridge University Press.

Hyland, Ken, 1996, "Writing Without Conviction? Hedging in Science Research Articles". *Applied Linguistics* 17/4: 433-454.

Hyland, Ken, 1998, *Hedging in Scientific Research Articles*, Amsterdam / Philadelphia, John Benjamins.

Kourilová, Magda, 1994, "Modal Meanings in Medical Science Reporting". In: Scott, Wendy / Mülhaus, Susanne (eds.), *Languages for Specific Purposes*, Kingston upon Thames, CILT and Kingston University: 65-67.

Koutsantoni, Dimitra, 2004, "Attitude, certainty and allusions to common knowledge in scientific research articles". *Journal of English for Academic Purposes* 3: 163-182.

Krug, Manfred, 2000, *Emerging English Modals: a Corpus-based Study of Grammaticalization*, Berlin / New York, Mouton de Gruyter.

Leech, Geoffrey / Rayson, Paul / Wilson, Andrew, 2001, *Word Frequencies in Written and Spoken English*, London, Longman.

Marqués Aguado, Teresa, 2008, "A Lexical Analysis of Modal Expressions of Certainty in Scientific Discourse". In: Lorenzo Modia, María Jesús (ed.), *Proceedings from the 31st AEDEAN Conference*, A Coruña, Universidade da Coruña: 373-385.

Miranda García, Antonio / Calle Martín, Javier / Moreno Olalla, David, 2001, *English for Science and Technology: Notional Description and Writing Practice*, Málaga, Servicio de Publicaciones de la Universidad de Málaga.

Palmer, Francis Robert, 2001, *Mood and Modality*, Cambridge / New York, Cambridge University Press.

Palmer, Francis Robert, 2003, "Modality in English: Theoretical, Descriptive and Typological Issues". In: Fachinetti, Roberta / Krug, Manfred / Palmer, Francis (eds.), *Modality in Contemporary English*, Berlin, Mouton de Gruyter: 1-17.

Pérez-Llantada Auría, Mª Carmen, 2008, "Stance and Academic Promotionalism: a Cross-Disciplinary Comparison in the Soft Sciences". *Atlantis* 30/1: 129-145.

Perkins, Michael R., 1983, *Modal Expressions in English*, London, Frances Pinter.

Pinar Sanz, María Jesús, 2005, "Review of *Perspectives on Evidentiality and Modality*", ed. by Juana I. Martín Arrese. *Atlantis* 27/2: 233-239.

Quirk, Randolph / Greenbaum, Sidney / Leech, Geoffrey / Svartvik, Jan, [1985] 2003, *A Comprehensive Grammar of the English Language*, Essex, Longman.

Scott, Mike, 1998, *WordSmith Tools Version 3.0,* Oxford, Oxford University Press.

Serrano Blanca, Melchor, 2006, *Creación de una base de datos léxica de términos utilizados en la ingeniería mecánica civil, relacionadas con las estructuras metálicas y de hormigón* (end-of-year project), Málaga, University of Málaga.

Sinclair, John, 1991, *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.

Swales, John M., 1990, *Genre Analysis: English in Academic and Research Settings*, Cambridge, Cambridge University Press.

Swales, John M., 2004, *Research Genres. Explorations and Applications*, Cambridge, Cambridge University Press.

Vihla, Minna, 2000, "Modal Expressions: Modifying Information". *Neuphilologische Mitteilungen* 101: 600-603.

Ziegeler, Debra, 2003, "On the Generic Origins of Modality in English". In: Hart, David (ed.), *English Modality in Context. Diachronic Perspectives*, Bern, Peter Lang: 33-69.